

Claims:

1 1. A system for parsing a piece of foreign language text into one or more phrases
2 which characterize a foreign language document, the system comprising:
3 a buffer for reading one or more words from the piece of text into the buffer until a break
4 character is identified;
5 a parser for identifying a phrase contained in the buffer, the phrase being a sequence of
6 two or more words in between break characters;
7 the parser further comprising means for determining the type of break character that
8 follows the identified phrase and means for saving a key phrase from the buffer based on the
9 determined type of break character ;
10 a database for storing the key foreign language phrases.

11 2. The system of Claim 1, wherein the buffer further comprises means for flushing
12 the buffer when the key phrase is stored in the database or the phrase in the buffer is deleted.

13 3. The system of Claim 1 further comprising a retriever for retrieving all occurrences
14 of the extracted phrases from the piece of text after the piece of text has been parsed.

15 4. A method for parsing a piece of text into one or more phrases which characterize
16 the document, the method comprising:
17 reading one or more words from the piece of text into a buffer until a break character is
18 identified;
19 identifying a phrase contained in the buffer, the phrase being a sequence of two or more
20 words in between break characters;

determining the type of break character that follows the identified phrase; and
 saving a key phrase from the buffer into a database based on the determined type of break
 character.

5. The method of Claim 4 further comprising flushing the buffer when the key
 phrase is stored in the database or the phrase in the buffer is deleted.

6. The method of Claim 4 further comprising retrieving all occurrences of the
 extracted phrases from the piece of text after the piece of text has been parsed.

7. A system for parsing a piece of text into one or more phrases which characterize a
 document, the system comprising:

a buffer for reading one or more words from the piece of text into the buffer until a break
 character is identified;

a parser for identifying a phrase contained in the buffer, the phrase being a sequence of
 two or more words in between break characters;

the parser further comprising means for determining the type of break character that
 follows the identified phrase and means for saving a key phrase from the buffer based on the
 determined type of break character ;

a database for storing the key foreign language phrases; and

a retriever for retrieving all occurrences of the extracted phrases from the piece of text
 after the piece of text has been parsed.

8. The system of Claim 7, wherein the buffer further comprises means for flushing
 the buffer when the key phrase is stored in the database or the phrase in the buffer is deleted.

9. A method for parsing a piece of text into one or more phrases which characterize the document, the method comprising:

- reading one or more words from the piece of text into a buffer until a break character is identified;
- identifying a phrase contained in the buffer, the phrase being a sequence of two or more words in between break characters;
- determining the type of break character that follows the identified phrase;
- saving a key phrase from the buffer into a database based on the determined type of break character; and
- retrieving all occurrences of the extracted phrases from the piece of text after the piece of text has been parsed.

10. The method of Claim 9 further comprising flushing the buffer when the key phrase is stored in the database or the phrase in the buffer is deleted.

11. A system for parsing a piece of text into one or more phrases which characterize the document, the system comprising:

- a first pass comprising means for identifying a phrase contained in a buffer wherein the phrase is a sequence of two or more words in between break characters, means for determining the type of break character that follows the identified phrase and means for saving a key phrase from the buffer based on the determined type of break character; and

7 a second pass comprising means for retrieving all occurrences of the extracted phrases
8 from the piece of text.

1 ~~12.~~ A method for parsing a piece of text into one or more phrases which characterize
2 the document, the method comprising:

3 performing a first pass through the piece of text, the first pass comprising identifying a
4 phrase contained in a buffer wherein the phrase is a sequence of two or more words in between
5 break characters, determining the type of break character that follows the identified phrase and
6 saving a key phrase from the buffer based on the determined type of break character; and

7 performing a second pass through the piece of text comprising retrieving all occurrences
8 of the extracted phrases from the piece of text.